

LABORATOŘ INTELIGENTNÍCH KOMUNIKAČNÍCH SYSTÉMŮ

—

KATEDRA INFORMATIKY A VÝPOČETNÍ TECHNIKY
FAKULTY APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY

Transkripce spontánní řeči

(manuál pro interní potřebu)

Kapitola 1

Transkripce zvukového záznamu

1.1 Potřebné vybavení

Transkripce zvukového záznamu není při současném stavu výpočetní techniky nijak výpočetně náročná činnost. Doporučený počítač musí mít takové parametry, aby byl schopen bez problémů a dostatečně svižně provozovat některý moderní operační systém, tj. například **Linux** či **Microsoft Windows 2000/XP**. Musí být pochopitelně vybaven zvukovou kartou schopnou přehrávat zvukový záznam vzorkovaný s frekvencí 44,1 kHz s rozlišením 16 bitů na vzorek, což je dnes již ale naprosto běžné.

K transkripci dále potřebujeme sluchátka. Záznam při transkripci totiž posloucháme zásadně kvalitními sluchátky s vyrovnanou frekvenční charakteristikou. Nejlepší jsou velké, tzv. *cirkumaurální* modely, u kterých náušníky obklopují celý ušní boltec a pomocí měkké obruby ho dokonale uzavírají. Je vhodné, aby se jednalo o **uzavřenou** konstrukci mušlí, protože ta nejlépe brání pronikání nežádoucích zvuků z okolí. Poslech pomocí reproduktorů nelze doporučit, protože reproduktory (zvláště levné modely k počítačům) zkreslují spektrum zaznamenaného signálu¹ a mohou některé části promluvy učinit nesrozumitelnými. Poslech doporučenými sluchátky (viz tabulka) zaručuje nejlepší možnou srozumitelnost. Poslech transkribovaného záznamu miniaturními/stylovými sluchátky od discmanů, MP3 přehrávačů, apod. nelze doporučit – jejich přednes je uzpůsoben reprodukci hudby (často ještě specifické).

¹Např. nežádoucím potlačením frekvencí ve střední části spektra a zesílením basových frekvencí – to je dnes typická konstrukční charakteristika běžně dostupných reproduktorů, protože nejlépe vyhovuje reprodukci moderní hudby.

| Výrobce | Model | Poznámka |
|------------|------------|---|
| AKG | K 271 | vynikající , mimořádně vyrovnaná frekvenční charakteristika, vysoká cena |
| AKG | K 100 | vyrovnaná charakteristika, dobrý poměr cena/výkon |
| AKG | K 66 | |
| KOSS | UR29 | dobry poměr cena/výkon |
| KOSS | R80 | vyrovnaná charakteristika, slušný útlum okolního hluku |
| Sennheiser | HD 250-II | vynikající , mimořádně vyrovnaná frekvenční charakteristika, vysoká cena |
| Sennheiser | HD 280 PRO | vynikající , excelentní útlum okolního hluku (až 32 dB) |
| Sennheiser | HD 215 | velmi čistý přednes, slabší útlum okolního hluku |
| Sennheiser | HD 270-V1 | |
| Sennheiser | HD 200-V1 | |

Tab. 1: Příklady vhodných sluchátek pro transkripci

1.2 Ergonomie a ochrana zdraví při transkripci

Vzhledem k tomu, že je transkripce časově velmi náročná činnost, je třeba dodržovat některé zásady, abychom se zbytečně neunavovali a nezpůsobovali si zdravotní komplikace.

Především myslíme na to, že strnulá poloha u počítače po dobu několika hodin bez pohnutí vede k úporným bolestem zad, krční páteře a následně k bolestem hlavy – v extrémních případech až k migréně. Při práci si proto stále „rovnáme“ záda, protahujeme se a pohybujeme hlavou – zejména do stran a dopředu. Hlavu při protahování nezakláníme; záklony ztuhnutí a bolest krční páteře naopak zesilují. Je vhodné minimálně jednou za hodinu práci na chvíli přerušit, udělat několik kroků po místnosti, napít se, protáhnout se, zakroutit trupem a pažemi, apod.

Dalším problémem je dlouhodobá fixace očí na monitor – když máme oči po dobu několika hodin zaostřené na stále stejnou vzdálenost na monitor, dochází k rychlé únavě zraku a zhoršuje se (často i trvale) vidění. Je proto třeba každou chvíli (minimálně jednou až dvakrát v průběhu 15 minut) odpoutat zrak od monitoru a přeostrit na „nekonečno“, například dlouhým pohledem z okna. Vhodné je také očima kroutit shora dolů, ze strany na stranu a dokola. Začnou-li se nám dělat při pohledu na monitor mžítka (obvykle tmavší pruhy nebo skvrny při pohledu na větší jednobarevnou oblast) je nejvyšší čas s prací pro dnešek skončit.

Poslech záznamu provádíme sluchátky – zde je další zdravotní riziko. Úroveň hlasitosti záznamu ve sluchátkách musíme mít nastavenou rozumně. Z tohoto důvodu jsou lepší uzavřené izolující modely, které brání vnikání nežádoucích zvuků zvenčí, a proto je možné je poslouchat při stejné srozumitelnosti na mnohem nižší hlasitost. Každopádně hlasitost záznamu nám musí být subjektivně příjemná. Pokud posloucháme delší dobu (což je při transkripci nevyhnutelné), hlasitost oproti subjektivně příjemné úrovni ještě o trochu snížíme. Máme-li při transkripci neustálý pocit, že záznam neslyšíme dobře a musíme přidávat hla-

sitost, znamená to, že je náš sluch již unavený a je třeba udělat alespoň čtvrt-hodinovou přestávku. Stálý dlouhodobý tlak náušníků sluchátek může způsobit bolest hlavy – je třeba používat sluchátka, která nepůsobí příliš velký tlak na hlavu a samozřejmě dělat přestávky.

Ve vlastním zájmu tyto zásady ochrany zdraví dodržujeme – pomáhají nám také předcházet nepozornosti, chybám a zkratům z únavy či bolesti hlavy, zad, apod.

1.3 Transkripční značky

(v následujícím textu je zachycený stav ke dni 1. prosince 2006)

V transkripci kromě běžného textu používáme ještě speciální značky, kterými se označují různé specifické zvuky, které je také třeba zaznamenat. Tabulka uvádí většinu používaných značek.

| Značka | Význam | Pův. anglický výraz |
|--------|--------------------------|-------------------------|
| +eh | ehmm/váhání | neutral vowel |
| +rs | dýchání | respiration |
| +is | nádech | inspiration |
| +es | výdech | expiration |
| +pu | funění | puffing |
| +bl | foukání | blowing |
| +sm | mlasknutí | smack |
| +xp | odkašlání | expectoration |
| +co | kašlání | cough |
| +la | smích mluvčího | laughter |
| +hs | syčení mluvčího | hiss |
| +kb | zvuk klávesnice | keyboard |
| +cl | zvuk myši | click |
| +pa | šustění papíru | paper rustling |
| +el | elektrický brum | electric noise |
| +mi | šum mikrofonu | microphone noise |
| +cv | konverzace v pozadí | background conversation |
| +bl | smích v pozadí | background laughter |
| +bc | kašel v pozadí | background cough |
| +tt | znělka | theme tune |
| +jg | jingle | jingle |
| +tp | top | top |
| +mu | hudba | music |
| +ap | potlesk | applause |
| +un | neidentifikovatelný zvuk | unidentified noise |
| +sl | ticho | silence |

Tab. 2: Transkripční značky

Značky se umísťují normálně do textu v místě výskytu daného zvuku, např.:

to je +eh dost +pu divné

Za značkou **musí vždy** následovat mezera, která značku oddělí od následujícího slova.

1.3.1 Transkripce koptání a nepravých začátků slov

V případě, že řečník začne slovo a nedokončí ho, načež totéž slovo řekne celé, píše se první – nedopovězená část (tzv. false start) – do závorek, tedy např.:

(ko)koptavý (ře)řečník

Takto ovšem označujeme pouze neúplná slova. Pokud řečník opakuje několikrát to samé slovo, ale vždy ho vysloví celé, zapíšeme ho normálně, např.:

jak jak jak se to to dělá

1.3.2 Problematická slova

Slova, která není možné transkribovat normálním způsobem, zapisujeme do dvojitých kulatých závorek, tedy např.:

((ukatval))

Mluví v tomto případě hodlal říct "ukazoval", ale uprostřed slova (asi) polkl, čímž vznikl výše uvedený patvar.

1.3.3 Zkratky a hláskování

Pokud mluví hláskuje nějakou zkratku (v uvedeném příkladu MHD), zapíšeme doslova to, co říká, tedy např.:

jel sem em há dě

Pokud ale mluví pronáší jen oddělená písmena (nehláskuje), vyslovuje vlastně vždy hlásku následovanou tzv. neutrální samohláskou neboli *schwa* (ta se běžně ve fonetické transkripci přepisuje symbolem ə). Schwa tedy musíme za vyslovené písmeno naznačit znakem -, protože změnění jeho výslovnost a zabrání např. asimilaci s následujícím slovem. Zápis tedy bude vypadat např.:

instrukce j- n- c- provádí

Kapitola 2

Transkripce v programu Transcriber

Program **Transcriber 1.5.1** je specializovaný software na pořizování transkripce zvukového záznamu pro účely rozpoznávání řeči. Na obr. 1 se můžeme seznámit s jeho podobou. V horní části (na obrázku šedá) je editor transkripce. Každá řádka (uvedená zelenou značkou) představuje transkripci jednoho segmentu signálu. Aktivní řádka, tj. ta kterou právě upravujeme, je označena světlejším pozadím.

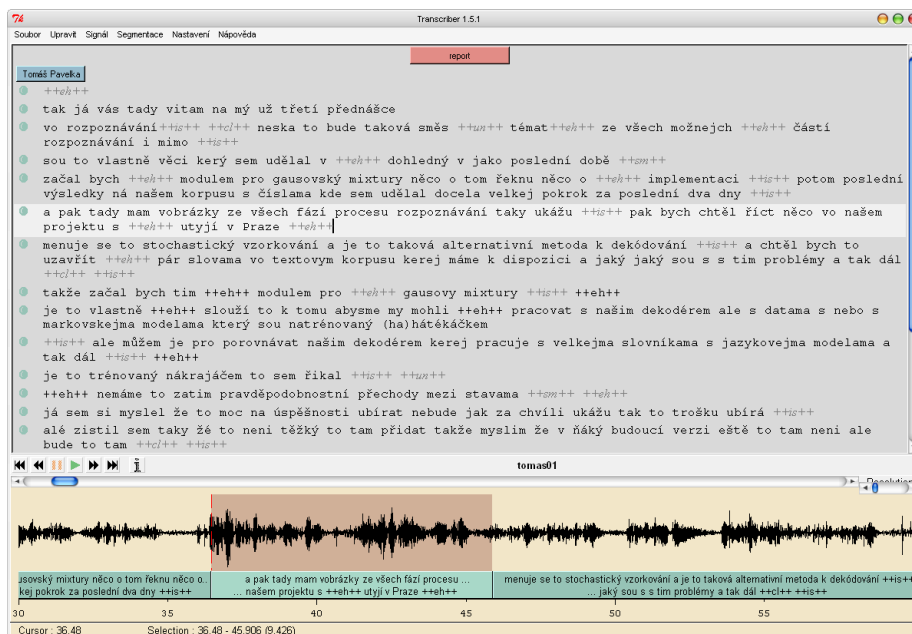
Ve střední části okna je ovládací panel přehrávání záznamu – tlačítka na něm zhruba odpovídají tlačítkům na CD přehrávači a stejný je i jejich význam. Zleva doprava je to přesun na předchozí segment, přesun o kus zpět, zastavení (pauza), přehrávání, přesun o kus vpřed, přesun na další segment.

Pod ovládacím panelem je zobrazen (zde černě na béžovém pozadí) zvukový signál záznamu. Posuvníkem nad signálem ho můžeme posouvat v okně tak, abychom viděli požadovaný úsek. Svislá přerušovaná červená čára značí pozici kurzoru, tj. které místo záznamu právě posloucháme. Stiskem levého tlačítka myši v oblasti zobrazení signálu můžeme kurzor přemísťovat a tak se po záznamu pohybovat.

V pravo nad signálem je krátký posuvník, kterým se nastavuje horizontální čili časové rozlišení signálu, tj. jak dlouhý úsek signálu se nám vejde do okna a tedy jak podrobné detaily signálu půjdou rozlišit.

Pod oblastí zobrazení signálu je pruh segmentace (zde v tyrkysové barvě). Pruh je svislými čarami rozdělen na segmenty. Hranice mezi segmenty se nazývá *breakpoint*. Každému segmentu odpovídá v editoru transkripce jedna řádka začínající zelenou značkou. Kliknutím levým tlačítkem myši si můžeme daný segment zvolit (např. k přehrávání) – současně se nám vyznačí odpovídající úsek zvukového signálu.

Pod segmentačním pruhem je ještě časová osa, na které jsou časové značky v sekundách od začátku záznamu. Dále je tam přesná pozice kurzoru v sekundách a ve stejných jednotkách začátek a konec vybraného úseku signálu.



Obr. 1: Program Transcriber 1.5.1

V dalších oddílech se dozvíte, jak postupovat k úspěšné transkripci zvukového záznamu.

2.1 Zahájení transkripce

Chceme-li v programu **Transcriber 1.5.1** začít vytvářet novou transkripci zvukového záznamu, vybereme z menu položku **Soubor/Nový transkript** (angl. **File/New trans**). Program se následně zeptá, který zvukový záznam (typicky soubor ve formátu WAV) chceme transkribovat. Vybraný zvukový soubor lze později změnit volbou z menu **Soubor/Otevřít zvukový soubor** (angl. **File/Open audio file**).

Mezivýsledky práce je třeba **často ukládat** volbou z menu **Soubor/Uložit** (angl. **File/Save**) nebo stiskem kombinace kláves **Ctrl+S**. Když ukládáme novou transkripci poprvé, program navrhne jméno výsledného souboru – stejné jako jméno transkribovaného zvukového souboru, jen s jinou příponou podle zvoleného formátu výsledného souboru (příponu nemusíme psát ručně, program ji sám doplní).

Chceme-li dále pracovat na již vytvořené transkripci, otevřeme ji volbou z menu **Soubor/Otevřít transkript** (angl. **File/Open trans**). Program vyhledá příslušný zvukový soubor se stejným jménem (ale jinou příponou), jako má soubor s transkripcí. Pokud takový soubor nenajde, zeptá se, kde je zvukový záznam příslušející k této transkripci na disku umístěn.

Pokud před ukončením programu uložíme nastavení volbou z menu **Nastavení/Uložit nastavení** (angl. **Options/Save configuration**), otevře se při příštím spuštění programu automaticky jak zvukový soubor, tak odpovídající soubor s transkripcí.

2.2 Poslech zvukového záznamu

Se zobrazeným signálem můžeme pracovat myší (přímo pomocí tlačítek nebo rolovacích lišt), výběrem z menu nebo pomocí klávesových zkratk¹. Chceme-li spustit nebo zastavit přehrávání záznamu, stiskneme klávesu **Tab**.

Při transkripci děláme nejdéle po 2 hodinách práce alespoň 15-ti minutovou přestávku, protože sluch se unavuje: Unavenému sluchu se může jevit některý úsek promluvy jako nesrozumitelný, zatímco odpočatý ho bez problémů zvládne transkribovat.

2.3 1. kolo: Hrubá segmentace

Zvolíme rozlišení signálu zhruba 30 sekund (nebo méně), tj. posuvníkem označeným **Resolution** v pravé dolní části okna nastavíme takové zobrazení, že se do okna vejde 30 sekund záznamu. Stiskem klávesy **Tab** spustíme přehrávání od začátku záznamu (nebo od začátku zatím netranskribovaného úseku záznamu). Během přehrávání vkládáme stisky klávesy **Enter** značky konce segmentu, tzv. *breakpointy*. Breakpoint bychom měli do transkripce vložit vždy, když udělá řečník pauzu nebo když je to vhodné kvůli logickému rozčlenění záznamu². Po vložení breakpointu se v textu (v horní části okna) objeví nová řádka a segmentační pruh (dole pod zobrazením signálu) se v tom místě rozdělí na dva úseky. Po několika minutách hrubé segmentace bychom měli mít celý záznam rozdělený na segmenty o délce zhruba 10 až 30 sekund. **Neustále v průběhu segmentace práci ukládejte!**

2.4 2. kolo: Transkripce

Zvolíme rozlišení signálu (viz výše) zhruba 10 sekund (nebo méně). Přesuneme se na první netranskribovaný úsek záznamu – buď kliknutím na první řádku textu (nahore) nebo na začátek signálu (dole). Posloucháme segment volbou z menu **Signál/Přehrát segment** (angl. **Signal/Replay segment**) nebo stiskem **Shift+Tab**.

Segment posloucháme tolikrát, kolikrát potřebujeme, abychom ho byli schopni bezchybně přepsat. Pokud i po několikatém poslechu stále nemůžeme porozumět, je vhodné si udělat krátkou přestávku. Můžeme také požádat někoho jiného (pokud je po ruce), aby si úsek poslechl a řekl nám, co porozuměl.

V případě, že hranice segmentu (na segmentačním pruhu dole pod zobrazením signálu) neodpovídá přesně transkripci³, posuneme hranici segmentu podržením prostředního tlačítka myši a tažením myši v požadovaném směru. **Je velmi důležité, aby hranice segmentů přesně odpovídaly transkripci.**

Když máme segment přesně ohraničený a rozumíme promluvě, zapíšeme transkripci do řádky v horní části okna. Současně se transkripce objeví i v segmentačním pruhu pod signálem. Zapsanou transkripci můžeme samozřejmě upravit kdykoliv potřebujeme. Po zapsání transkripce je vhodný okamžik k **uložení práce** stiskem **Ctrl+S**.

¹tj. kombinací kláves

²Zde bohužel neexistuje nějaké přesné pravidlo – vyžaduje to trochu citu a zkušeností.

³Např. se může stát, že je poslední slovo segmentu useknuté nebo naopak do segmentu zasahuje začátek slova ze segmentu bezprostředně následujícího.

Opakováním výše uvedeného postupu transkribujeme všechny vytvořené segmenty záznamu.

2.5 3. kolo: Verifikace

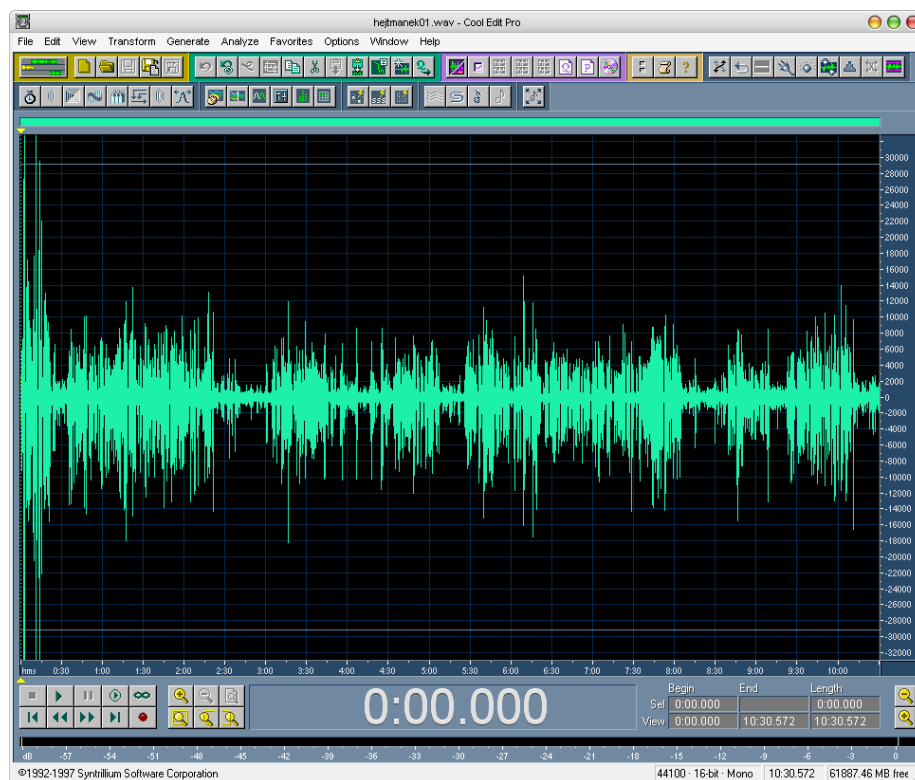
Vrátíme se na začátek záznamu, spustíme přehrávání, pečlivě posloucháme signál a současně čteme transkripci. Ověřujeme také, že hranice segmentů odpovídají transkripci. Stiskem **Tab** si přehrávání podle potřeby pozastavujeme a opět spouštíme. Je také možné během přehrávání přeskokovat po záznamu vpřed nebo zpět – myší, stiskem tlačítek na ovládacím panelu přehrávání nebo stiskem **Alt+šipka vpravo/vlevo**.

K úplnému a spolehlivému ověření přesnosti transkripce obvykle nestačí pouze jeden poslech. Chyby v transkripci, na které narazíme, opravujeme rovnou, abychom na ně nezapomněli. Každopádně během posledního kontrolního poslechu už bychom žádnou chybu odhalit neměli. A tedy opačně: Narazíme-li na chybu, musíme provést po jejím opravením další kontrolní poslech. **Průběžně práci ukládáme.**

Kapitola 3

Úpravy zvukového signálu

Před samotným započítím transkripce je vhodné provést některé úpravy zvukového záznamu.



Zvukový editor Cool Edit Pro